What is a Beryllium Measurement?
A Critical Look at Beryllium Quantitation

Charles Davis[1], Dan Field[2], John Hess[3] and Dan Jensen[4]

Abstract and Summary

DOE workplaces strive to comply with the 10 CFR 850.31(b)(1) surface concentration release criterion. The usual planning considerations for demonstrating compliance are these: how many swipes, and where; which sample preparation and analytical methods; what reporting limits; and what sample statistic to compare with the criterion.

We have reviewed swipe samples from hundreds of Nevada Test Site workplaces: office buildings; experimental facilities; forward area field units; shops; and tunnels. Our experiences have led us to a critical examination of the inner workings of the measurement process itself, involving details generally taken for granted when those usual questions are asked.

In this presentation we dissect the ICP-AES Be measurement process. We discuss calibration options and how they impact the distributions of analytical results. We look at distributions of blank results obtained from different labs, and discuss their relevance to determining reporting limits. We examine the way measurements are made from spectra, how that process impacts our understanding of the actual statistical distributions of Be measurements, and how interferences can affect Be measurements. Our objective is to gain sufficient confidence in the measurement process so that the usual questions will make sense and the survey results will be credible.

Based on our observations, we offer these recommendations:

- prepare calibration samples in digested blank swipes;
- force the calibration line through (0,0);
- base reporting limits on field blank measurement distributions rather than 40 CFR 236 calculations;
- use, but do not believe, the usual lognormal distribution assumption; and
- avoid the 234.861 nm emission line.

1 Principal Statistician, EnviroStat (charles.davis@envirostat-nv.com)
2 Industrial Hygiene Program Manager, Bechtel Nevada (fieldde@nv.doe.gov)
3 Senior Scientist, Bechtel Nevada (hessjw@nv.doe.gov)
4 Industrial Hygienist, Fluor Hanford (Daniel_A_Jensen@rl.gov)

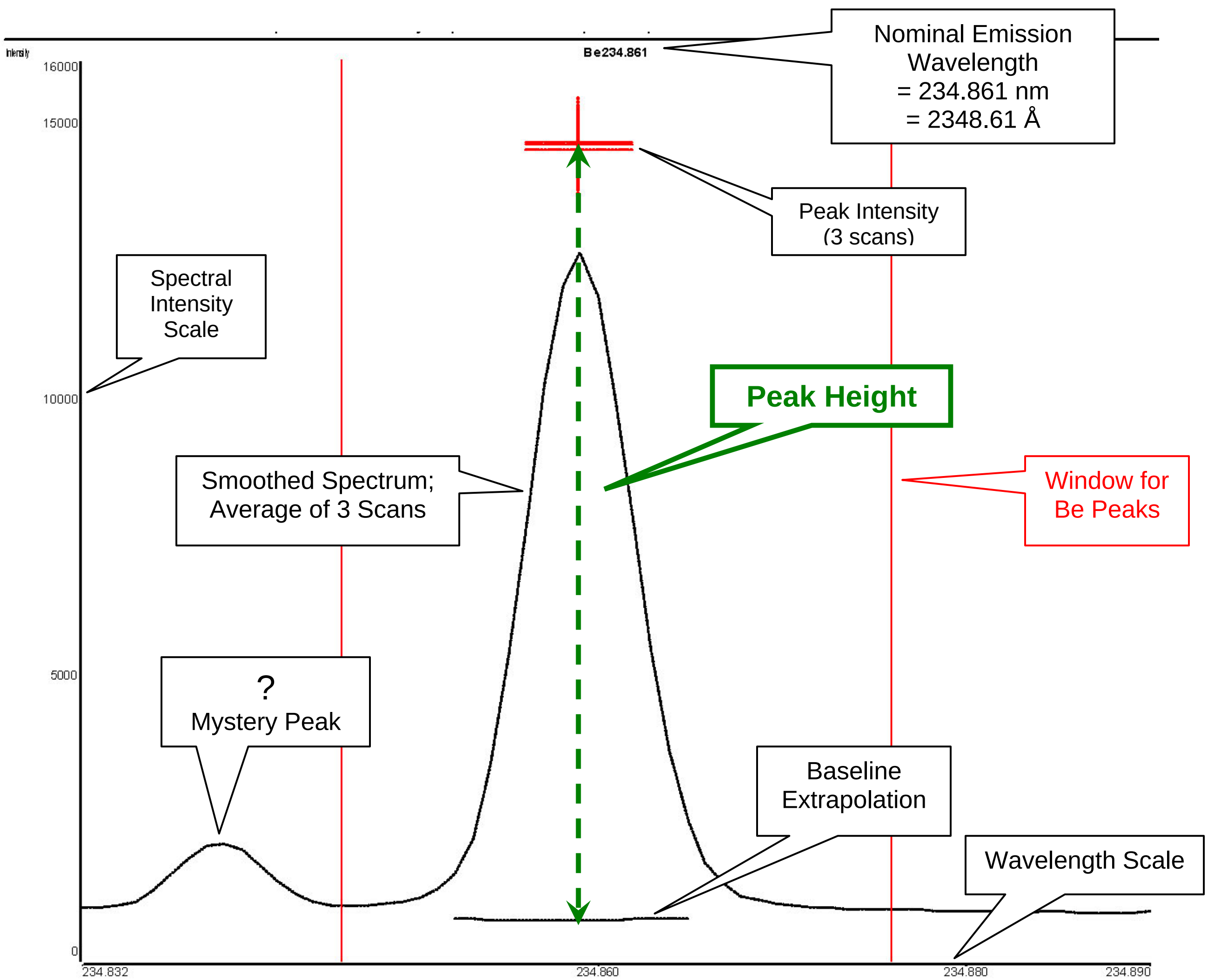
The Anatomy of a Beryllium ICP-AES Measurement

The Be ICP-AES measurement process involves four components:

- the sample;
- the measurement of its emission spectrum;
- the quantitation made from that spectrum and a calibration line; and
- the distributions of quantitated measurements of blanks and non-blanks.

The matters of planning sampling campaigns, obtaining samples, and interpreting the resulting data have been discussed at length elsewhere. We concentrate on the remaining three components.

The Spectrum



This is a Ghost Wipe spiked with 2.0 µg Be; the measurement was 1.859 µg/swipe.

Points worth noting

Spectrum: Three scans are made of each sample. Each scan is quantitated separately, and the three results are averaged. The spectrum shown is a smoothed version of the average spectrum at each wavelength. The actual **peak intensity** (PI) used in quantitation is shown by the red **crosshairs**.

A critical point is that the emission "line" is actually a broad peak whose width increases as its height increases.

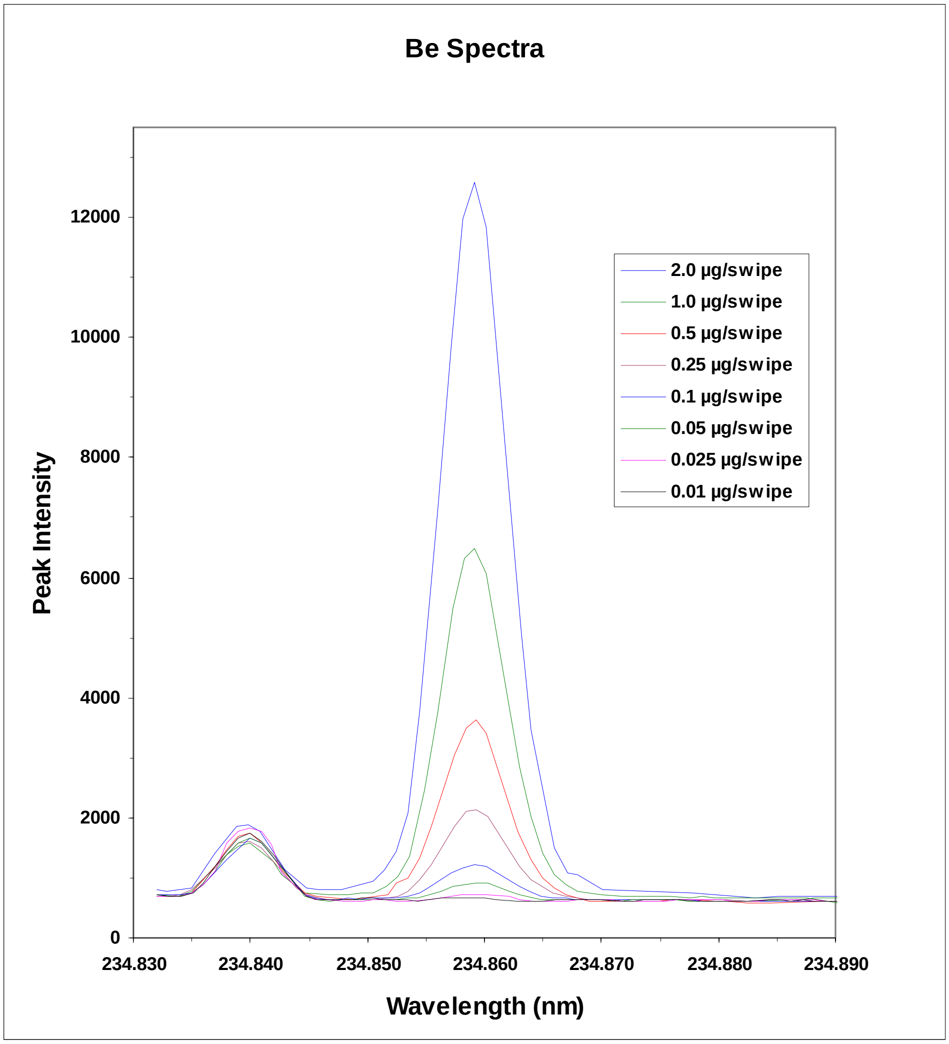
Nominal wavelength: Since instruments can drift, the algorithm will find a peak within the **window** to quantitate, and will expand the window if it needs to in order to find a peak.

Baseline: This is extrapolated from adjacent areas of the spectrum. The **peak height** (PH) is measured from the baseline to the PI on the actual spectrum, not the smoothed spectrum.

Mystery peak: This is an interferant with emission wavelength approximately 234.841 nm. It appears innocuous in the spectrum shown above, but this was of a spiked Ghost Wipe with no other matrix effects present! The following plots show that this peak, apparently from iron (Fe), is of sufficient magnitude to possibly affect quantitation of low levels of Be even if there are no other matrix effects present. In natural, and particularly industrial, settings Fe concentrations can be considerably higher than the trace amounts present in Ghost Wipes, of course.

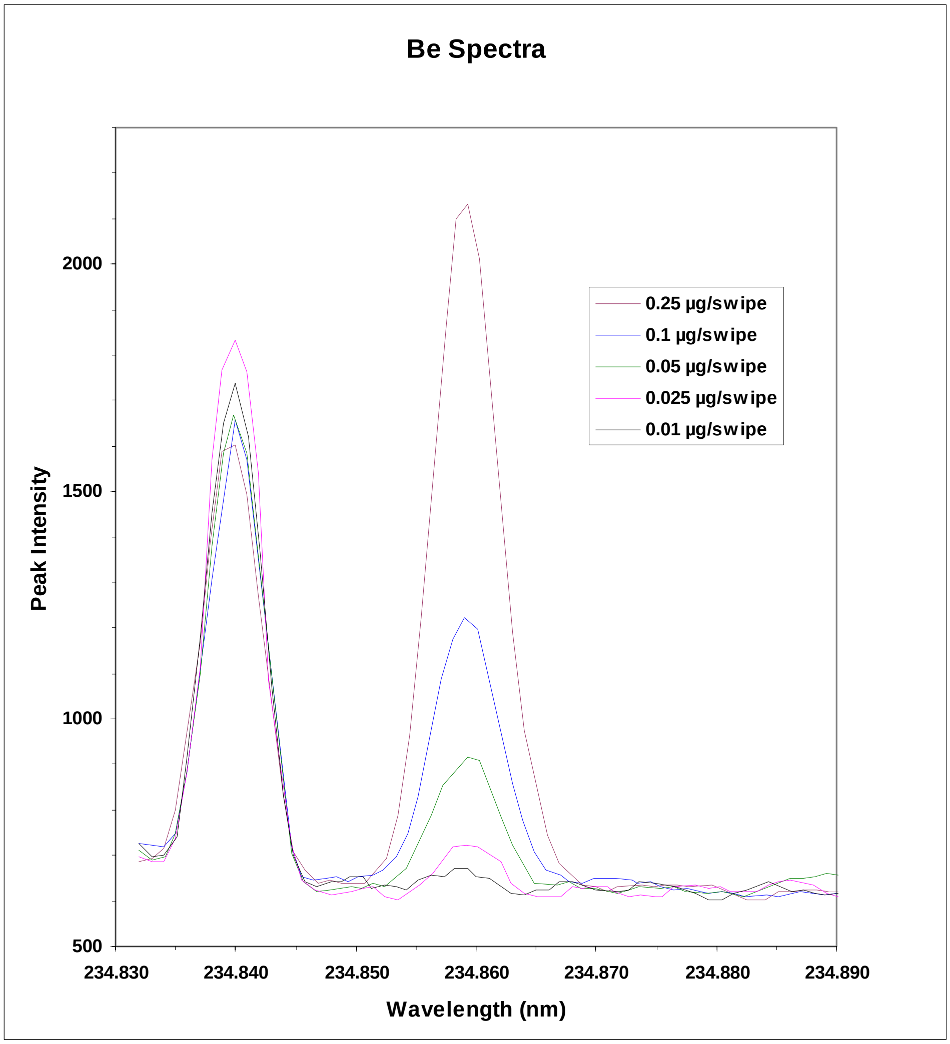
This interferant peak is apparently a rather minor emission peak of Fe; it is not listed in the *NIST Handbook of Basic Atomic Spectroscopic Data*, for instance; see http://physics.nist.gov/PhysRefData/Handbook/index.html.

Spectra for Various Spiking Concentrations



These are of spiked Ghost Wipes with "made-to" concentrations range from 2.0 µg/swipe to 0.01 µg/swipe. The baselines, actual PIs crosshairs, and windows are suppressed here.

The next plot shows the lower five concentrations of these eight.

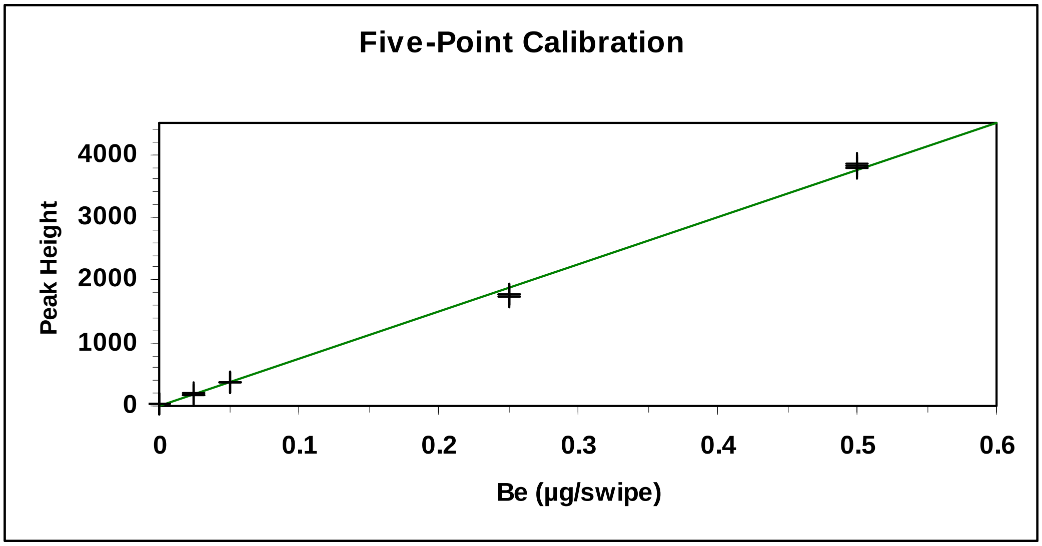


The "made-to" concentrations here range from 0.25 µg/swipe to 0.01 µg/swipe. This covers the range of concentrations of interest for many Be field investigations.

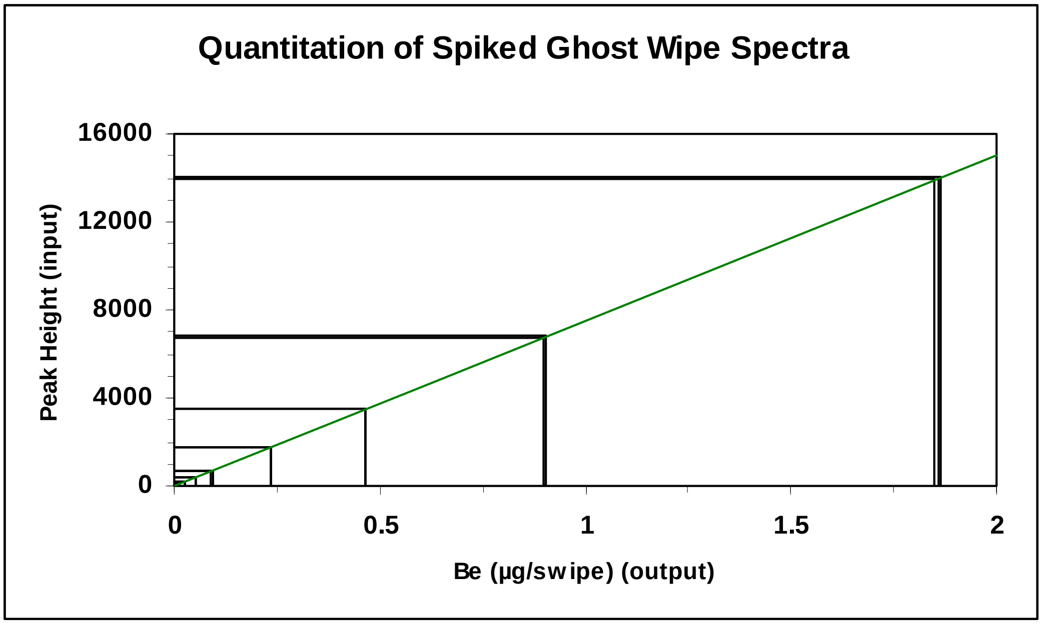
Note the relative sizes of the Be peak and the Fe peak at 234.841 nm. This Fe peak is present in spectra involving Ghost Wipes; there is no similar peak with spiked water samples. One implication of this is that it makes sense to prepare calibration samples in digested Ghost Wipes, if that is the sampling medium, even if a different nominal emission line is used. The Fe content of a Ghost Wipe is approximately 10 µg.

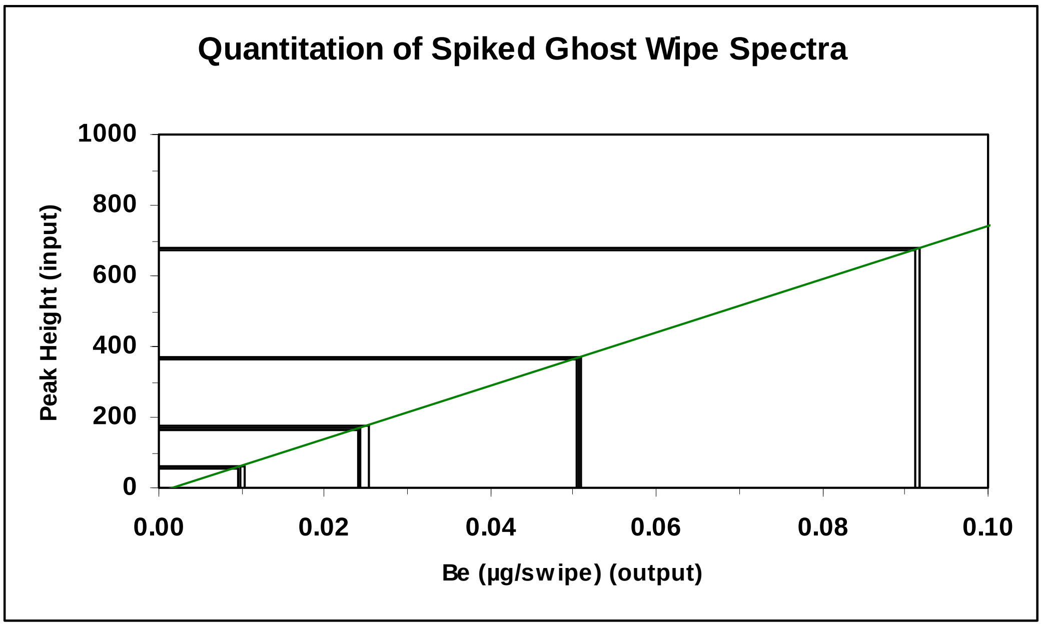
The Calibration Line

The calibration is based on five standards, with solution concentrations from 0 to 10 ppb, corresponding to swipe concentrations from 0.0 to 0.5 µg/swipe. There are three scans for each standard. In this calibration the relative standard deviations at the five concentrations range from 15.65% at the low end to 0.34% and 0.65% at the upper values. The calibration line is PH = -14.843 + 7542.7 * Be (in µg/swipe). For this calibration line, a PH of 0 would correspond to Be = 0.002 µg/swipe; the calibration line does not quite go through (0,0).



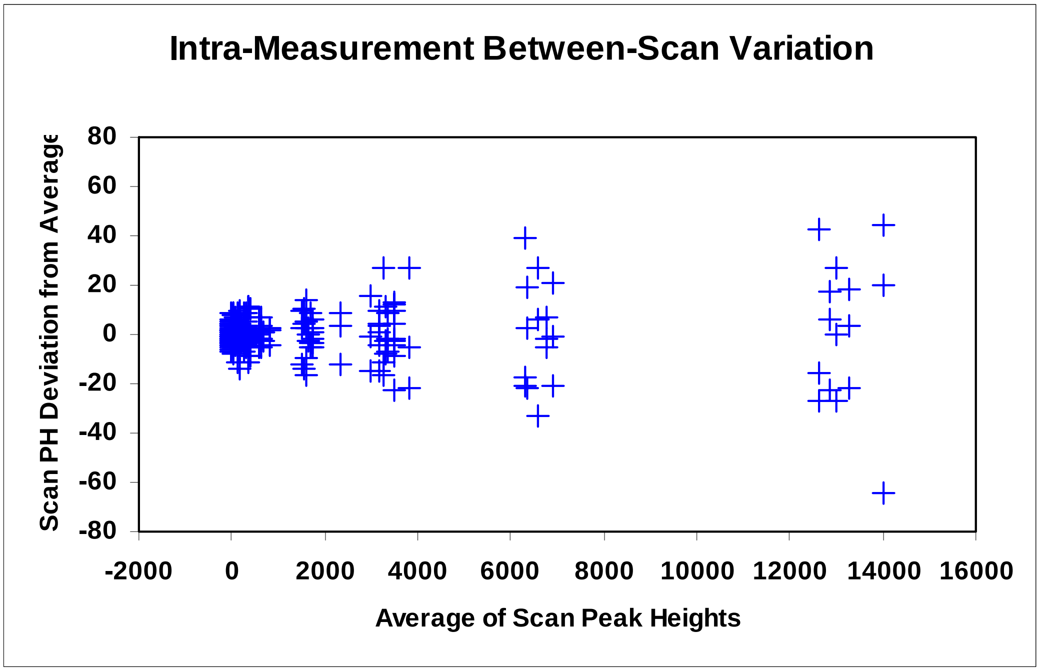
Then the PH is read from a spectrum and reflected through the calibration line to the Be scale. In each case there are three scans per measurement. The next two figures display the quantitations of the eight spectra for spiked Ghost Wipes shown above.



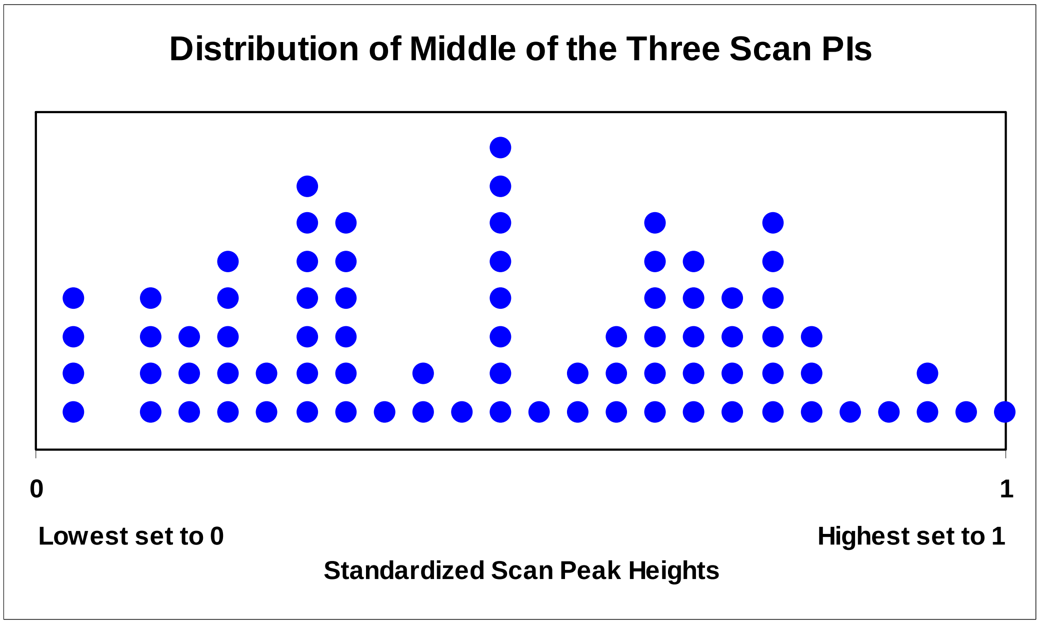


The second contains just the lower four concentrations; the first contains all eight. In these examples the three scans for each measurement agree quite nicely.

By the way, if we look at the variability among scans in a single quantitation, we get pictures like the following. The first plots the variability of the individual scan peak heights against their average, showing that measurement variability, even at the instrument level, clearly increases with increasing concentration.



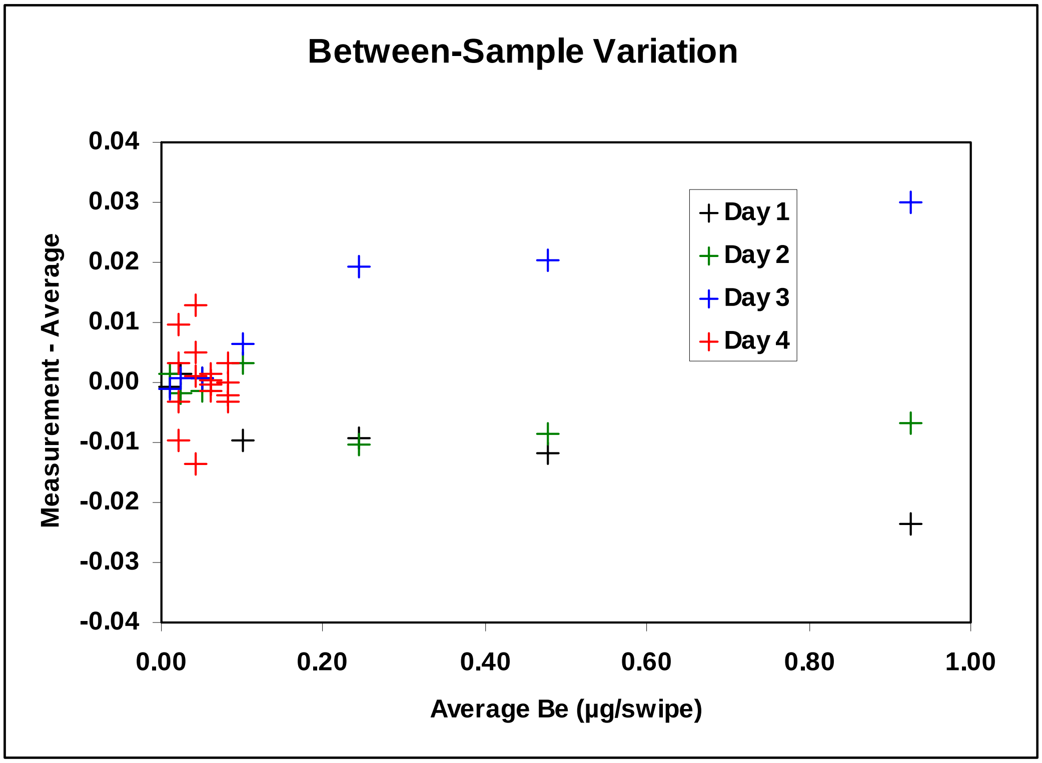
In the next we standardize by setting the lowest PH to 0 and the highest to 1, and plot the distribution of the middle of the three. This distribution is nicely symmetric; its coefficient of skewness is a mere 0.01. Its normal probability plot correlation coefficient is 0.984; it just barely fails the Ryan-Joiner probability plot correlation coefficient normality test (p = 0.047) because the short tails of the distribution and the large number of data points. We return to this thread when looking at distributions of blank measurements later.

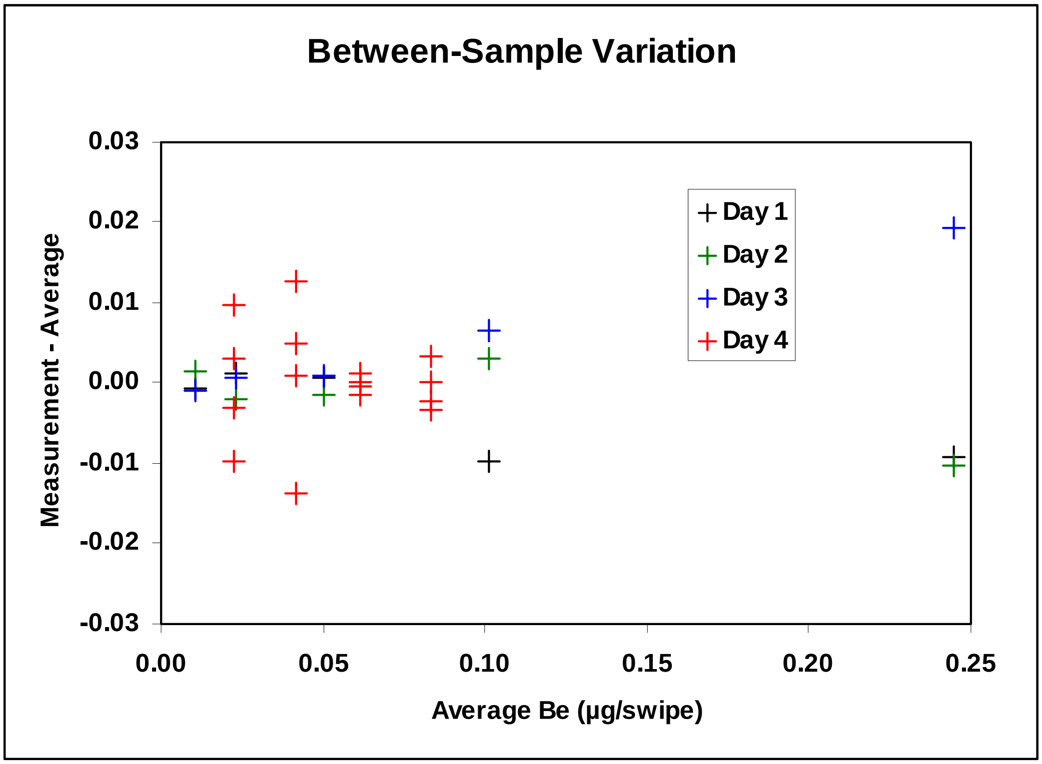


The preceding two plots show only instrument variation <u>within the same measurement</u>. Variation between measurements, even spikes made with the same concentrations and no matrix effects, will be greater. In addition to instrument variability we have differences in the specific Ghost Wipes, different calibrations, different digestions, instrument drift, and other factors.

The next two plots show the deviations from the average quantitations for spiked Ghost Wipes made to concentrations from 0.01 to 1.0 µg/swipe. These were run on different days; the instrument was disassembled and reassembled between days 3 and 4. The second plot is just the left end of the first plot.

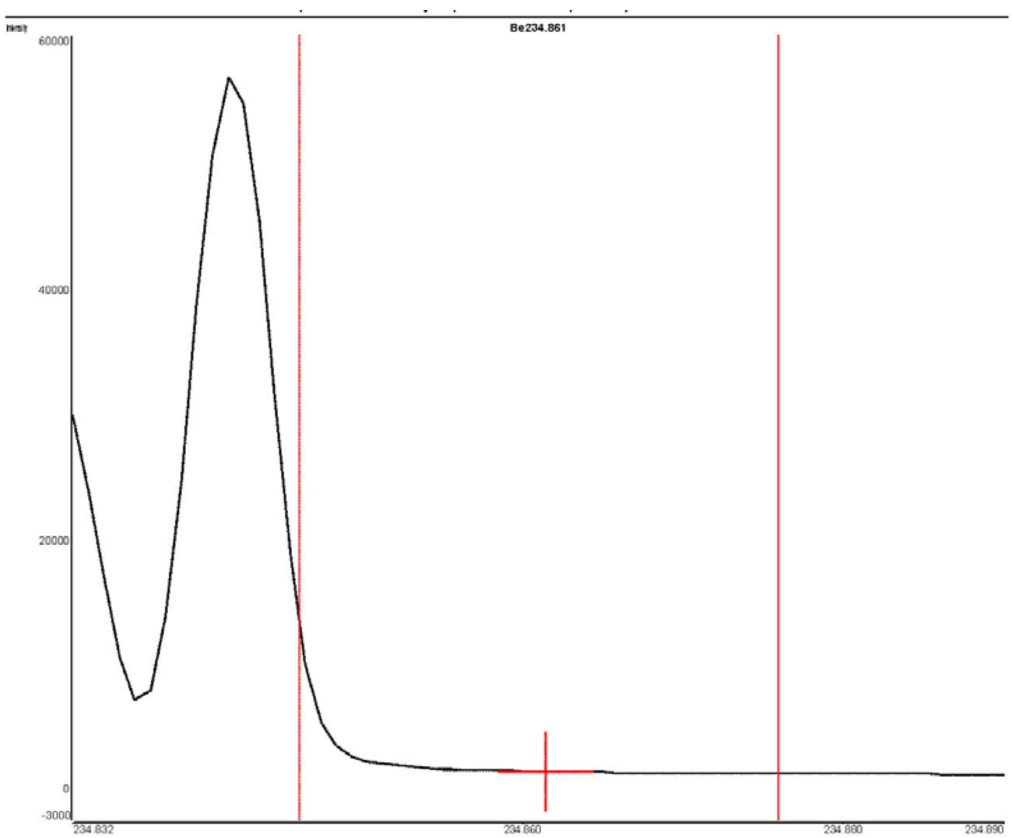
Real-world matrix effects would add even more variation, of course.


Between-Sample Variation
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
-0.04
Measurement - Average
0.00
0.20
0.40
0.60
0.80
1.00
Average Be (µg/swipe)
Day 1
Day 2
Day 3
Day 4


Between-Sample Variation
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
Measurement - Average
0.00
0.05
0.10
0.15
0.20
0.25
Average Be (µg/swipe)
Day 1
Day 2
Day 3
Day 4

Negative "Concentrations"

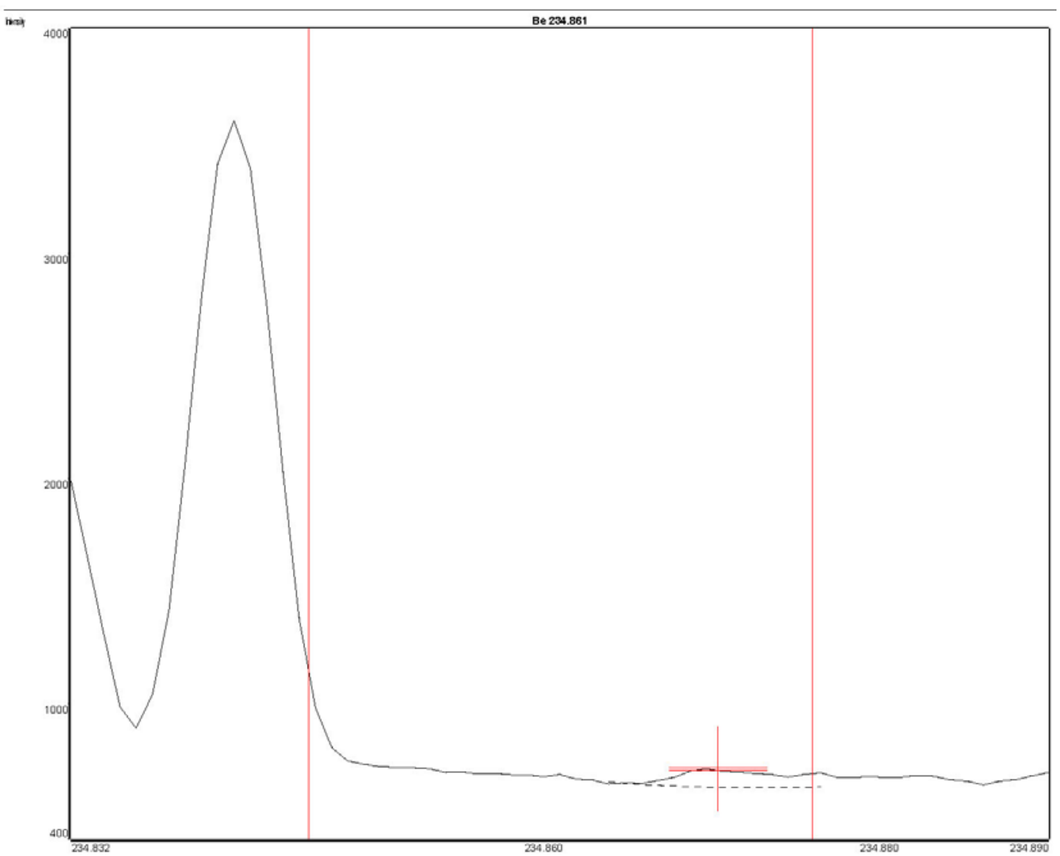
It can easily occur, and does occur a fairly large proportion of the time in some circumstances depending on the calibration details among other factors, that a quantitation results in a negative value. Here is one example; this is a Ghost Wipe spiked with 0.01 µg Be and 500 µg Fe. The PI scale extends from -3000 to 60,000 and the measurement was reported as -0.0047 µg/swipe for this sample. Our interpretation is that the baseline, which is extrapolated from the spectral regions away from 234.861 nm, is highly influenced by the tail of the Fe peak, resulting in a negative PH.



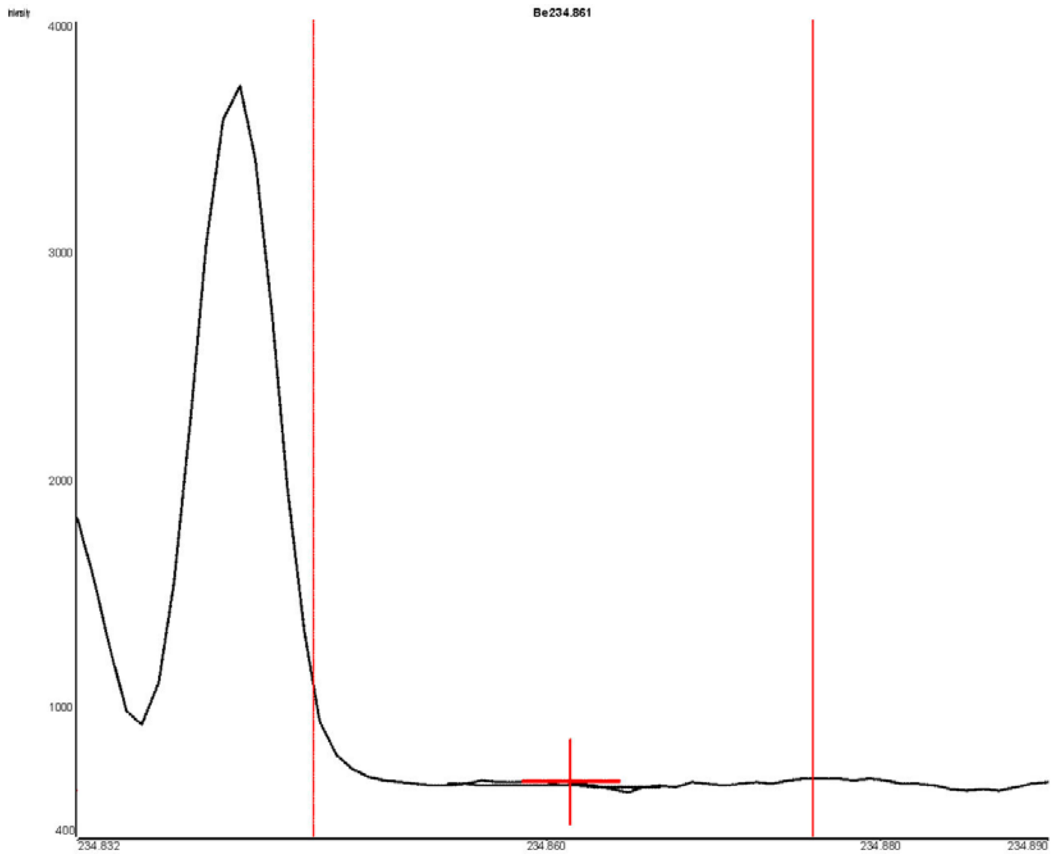
Real-life data from some of our facilities consisted of well over 50% negative values.

Real Samples

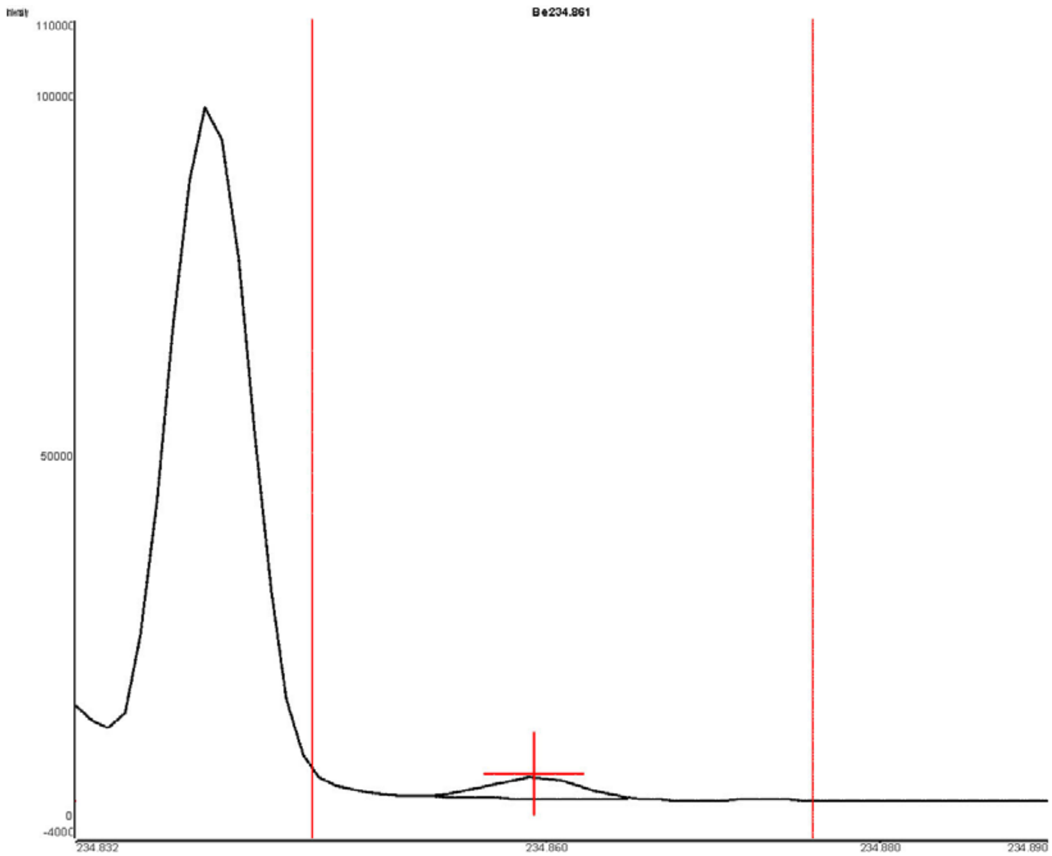
So what do spectra for real samples look like? Here's a regular uneventful measurement. This swipe was obtained from atop a fan in a dormitory area in Area 6 on the NTS. The measurement was 0.0129 µg/swipe. The PI scale ranges from 400 to 4000. Note the offset of the peak from the center of the window.



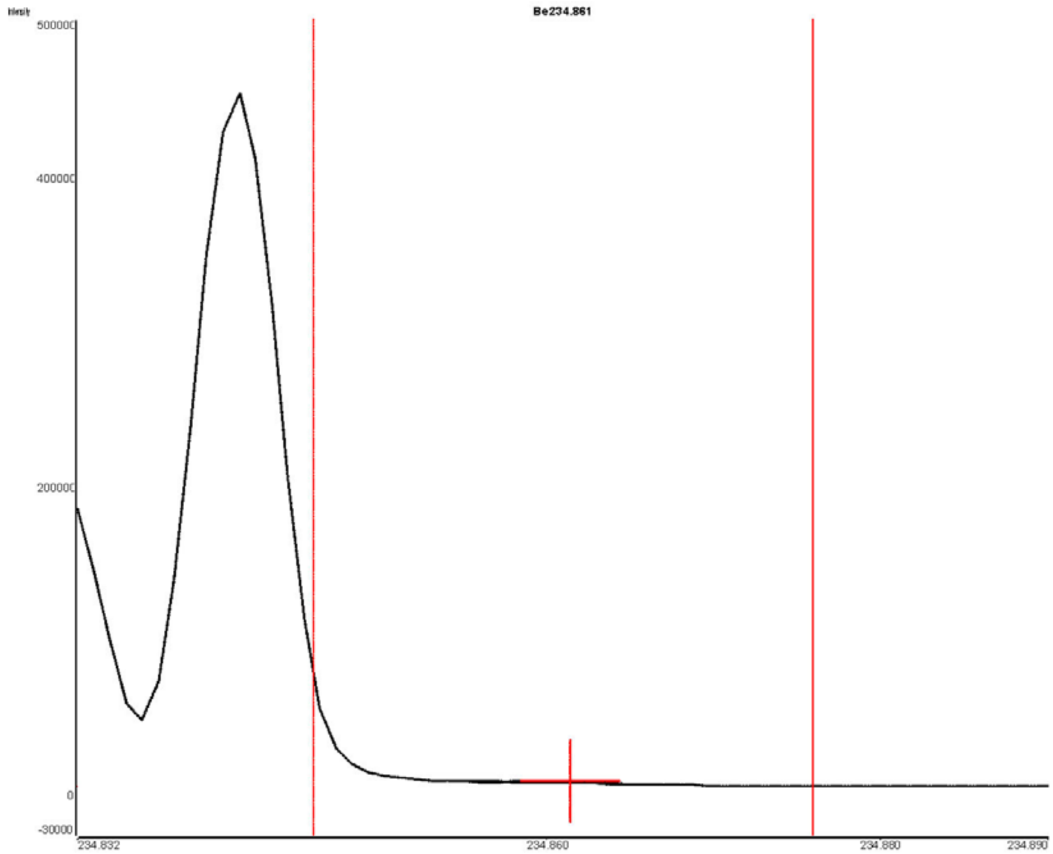
Here’s another regular uneventful sample, taken from a shelf in the communications room in the Mercury, NV Fire Department. The measured Be concentration is 0.004 µg/swipe; this was right around the <u>stated</u> LOD for this particular analysis. (Our sampling and analysis plans call for all observations to be quantitated, which has turned out to be highly educational.) The PI scale again goes from 400 to 4000.



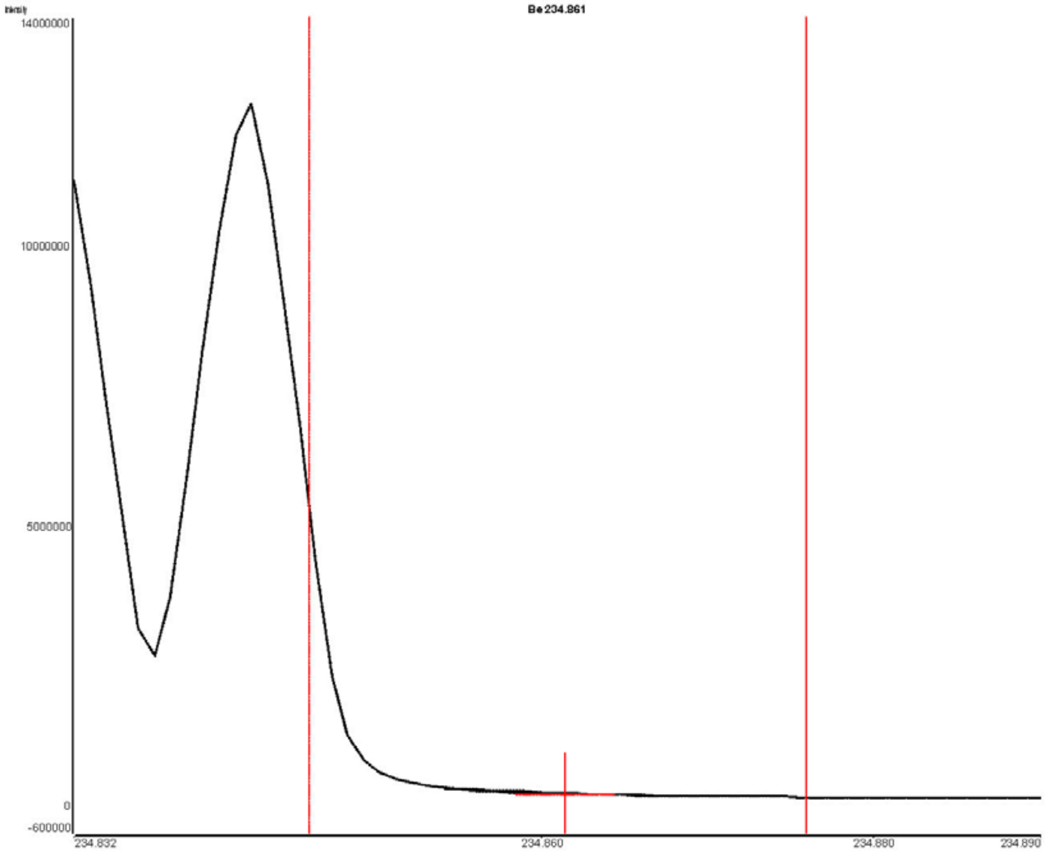
This sample was taken atop a dusty electrical box in a rarely-entered video projection room in a training facility in Mercury. The quantitation was 0.247 µg/swipe. The curiosity here is the very large PI scale, extending from -4000 to 110,000. There's a lot of Fe in that dust!



Here's another spectrum from a dusty environment, this time a trailer in a forward area. The quantitation was 0.110 µg/swipe; the PI scale extends from -30,000 to 500,000. The Be peak and baseline are almost negligible compared with the broad interferant peak.

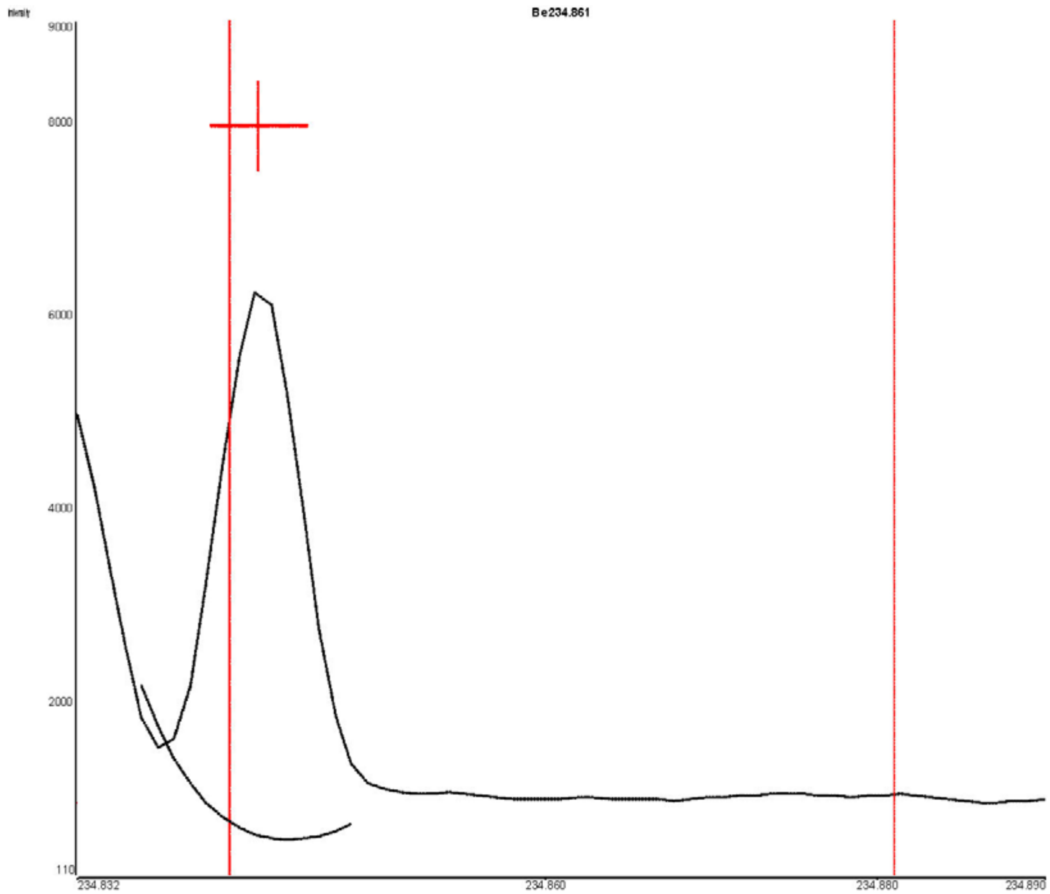


This sample was obtained on a band saw in a metal machining shop in NTS Area 6. The quantitation was a whopping -1.489 µg/swipe; the PI scale extends from -600,000 to 14,000,000.



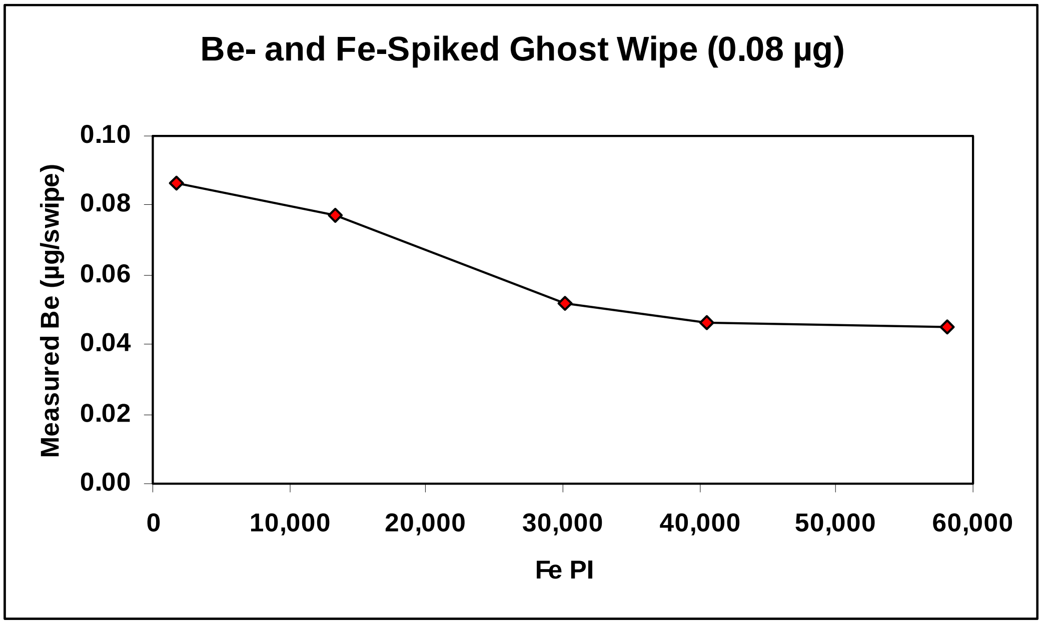
In cases like the last two, we believe that the baseline extrapolated from the neighboring spectrum is basically random noise, so that the resulting measurement values are unreliable.

This sample was taken from the top of a file cabinet in Los Alamos. The quantitation was 0.533 µg/swipe. Note, though, that the instrument's algorithm had to expand the window to find a peak to quantitate, and that it is highly probable that it found and used a slightly shifted Fe peak rather than Be! This result is surely spurious.



Interpretation

After seeing these results, and having a large number of samples that had been run using the 234.861 Be peak, we became highly motivated to understand the mechanics of the Fe interference better. Here are some further results using spiked Ghost Wipes. This particular set of five samples is spiked with 0.08 µg/swipe of Be and 0.0, 0.12, 0.26, 0.4, and 0.58 mg/swipe Fe. These results are typical; the Be quantitation is relatively unaffected for Be concentrations above around 0.1 µg/swipe. We believe that an interference correction is possible so long as the Fe PI is no greater than around 70,000. In that region a simple correction for measured Be values less than 0.1 will work adequately. Conversely, if the Fe PI is greater than 70,000, the measured Be values should not be relied upon.



Distributions of Blank Measurement Results

We now shift gears from examining spectra to examining the distributions of blank measurements. Here are plots showing results from three laboratories using four different measurement systems. The systems differ by swipe media, emission line used, and most likely calibration details as well. First, here are the numbers of observations and details of the measurement systems to the extent they are known.

Table 1
Numbers of Observations/Percent Censored

Medium	Lab A	Lab B	Lab D
Ghost Wipe	0	287/0.4%	76/12%
Whatman Filter	11/64%	2/0%	286/41%

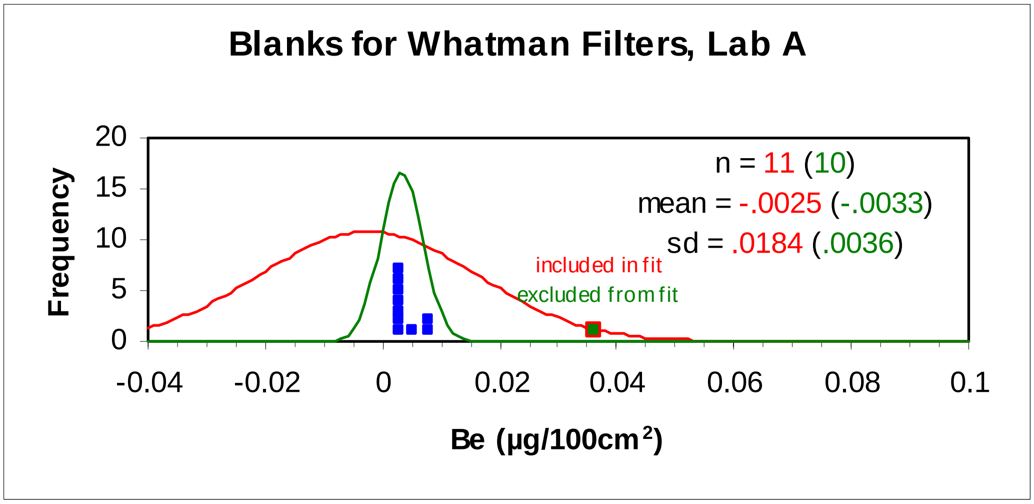
Table 2
Sample Preparation and Analysis

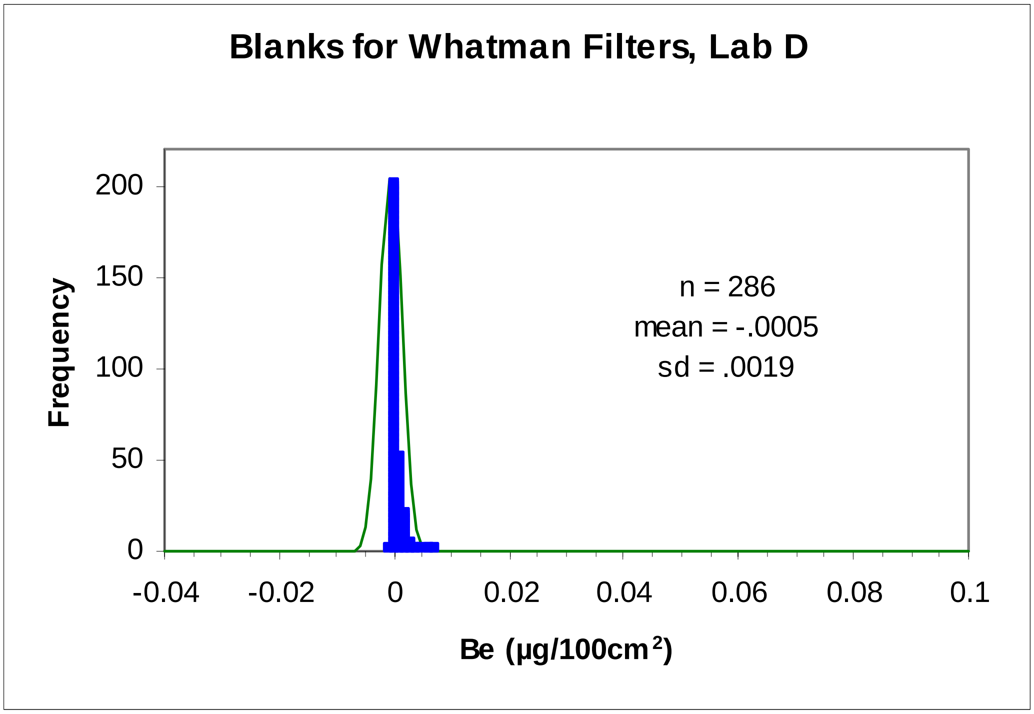
Lab	Stated Method(s)	Digestion	Blank Correction?	Line (nm)	Stated LODs
A	NIOSH 7300	H_2SO_4, HNO_3, HCl; $HClO_4$; heat	Yes	313.042?	0.005
B	NIOSH 7303 Modified	HNO_3, heat	No	234.861	0.003-0.005; sometimes 0.01
D	NIOSH 7303 Modified	HNO_3, heat	Yes	313.042	0.003-0.005; sometimes 0.02

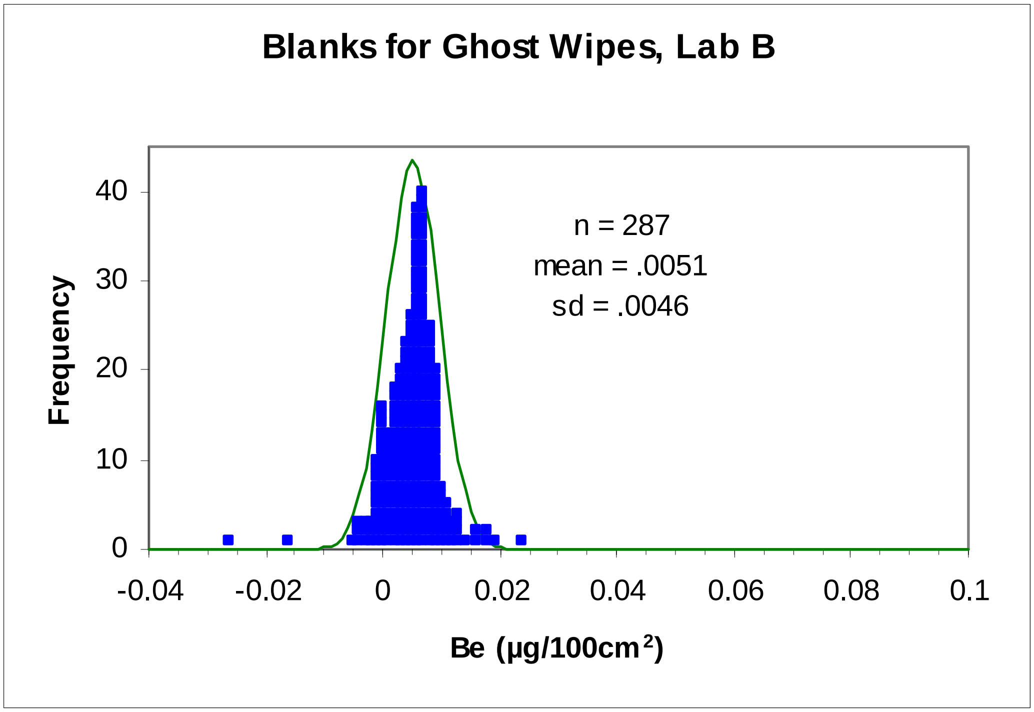
* The standard emission line for ICP-AES by Method 6010C is 313.042

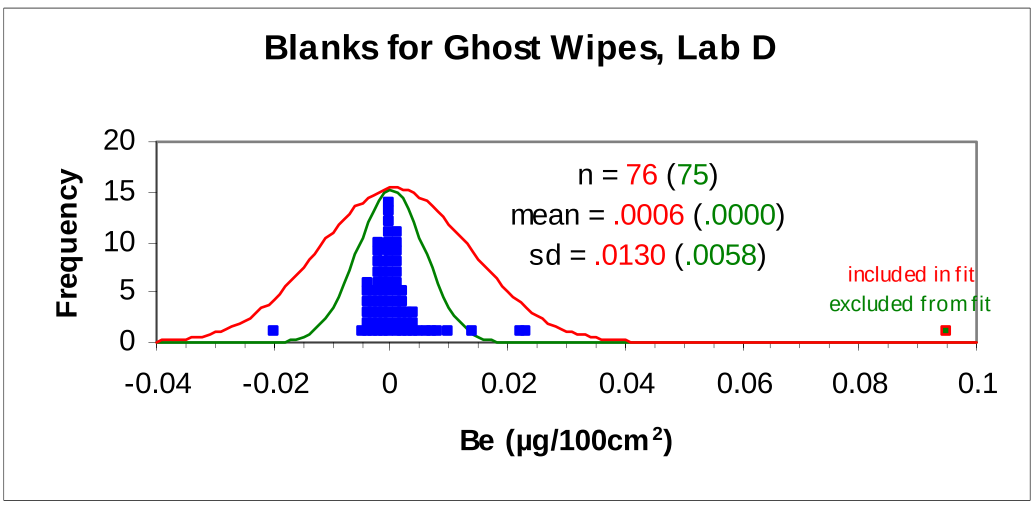
Some observations were left-censored; i.e. reported as “<0” (labs B and D) or “<0.005” (lab A). The proportion of censored data is included in Table 1. These censored data are plotted at their censoring points; this results in some apparent skewness in the distributions for Whatman Filter blanks. A censored data normal distribution maximum likelihood algorithm was used to estimate means and standard deviations.

Labs A and D blank-corrected their water-based calibration standards by forcing the calibration line though zero. Additional compensation for matrix effects like the Fe peak displayed in the previous scans can be obtained by using pre-digested media for calibration standards and blanks.


Blanks for Whatman Filters, Lab A
n = 11 (10)
mean = -.0025 (-.0033)
sd = .0184 (.0036)
included in fit
excluded from fit
Frequency
20
15
10
5
0
-0.04
-0.02
0
0.02
0.04
0.06
0.08
0.1
Be (µg/100cm²)


Blanks for Whatman Filters, Lab D
n = 286
mean = -.0005
sd = .0019
Frequency
200
150
100
50
0
-0.04
-0.02
0
0.02
0.04
0.06
0.08
0.1
Be (µg/100cm²)





The actual and fitted distributions for the blanks vary considerably among the measurement systems. The next plot shows the fitted distributions superimposed; the fits shown exclude the outliers noted in the plots.

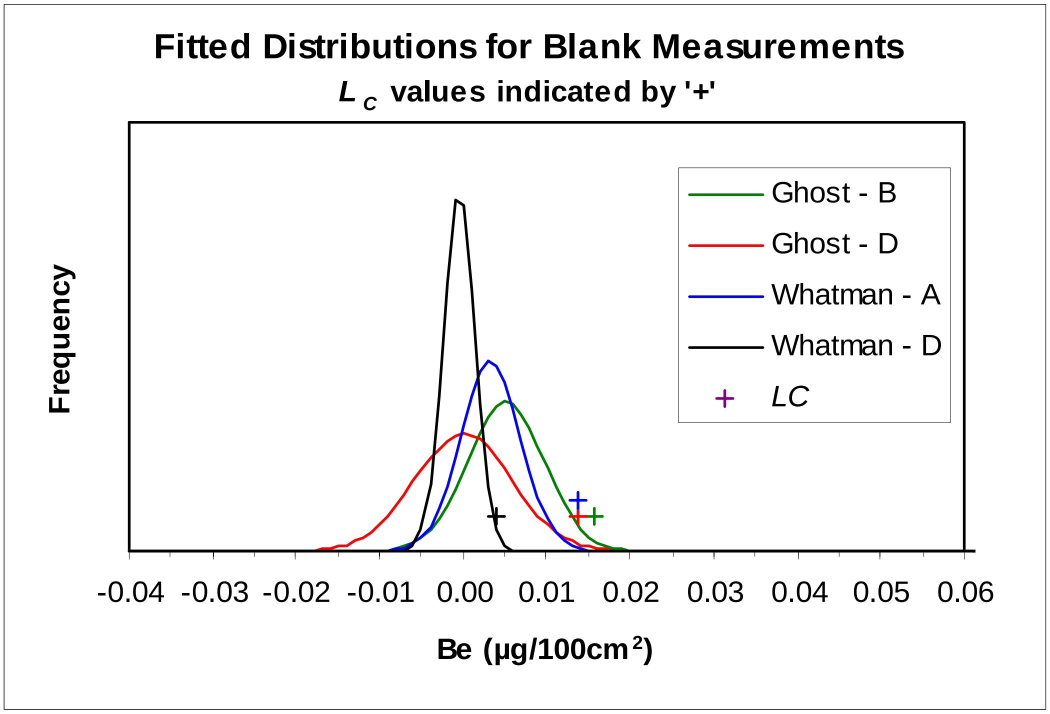


Table 3 summarizes the numerical results from the fitting.

Table 3
Fitted Means / Standard Deviations

Medium	Lab A	Lab B	Lab D
Ghost Wipe	N/A	0.005100 / 0.004584	0.000037 / 0.005799*
Whatman Filter	0.003256 / 0.003601**	N/A	-0.000538 / 0.001896

* 0.000639 / 0.012971 with high outlier included
** -0.002495 / 0.018412 with high probable outlier included

Empirical Critical Values (L_C)

Finally, Table 4 shows the empirical L_C values derived from these distributions, along with the typical LODs reported by the laboratories. The LOD computations are based on 40 CFR Part 136 Appendix B computations from spiked solutions. L_C, by contrast, is the critical value for a test of the hypothesis that a measurement is statistically significantly (1% significance level) higher than a blank measurement; see Currie (*Analytical Chemistry* 40, 586-593, 1968). The empirical L_C values are computed as upper 99% prediction limits of the distribution of blank measurements; this is equivalent to determining the lowest value that would be statistically significantly higher than a blank using a 1% significance level *t*- test.

The empirical L_C value for Whatman Filters for lab A appears higher relative to the fitted distribution than in the other cases, due to the small sample size (10 after outlier deletion). The L_C values are also shown in the previous plot.

Table 4
Fitted L_C Values / Typical Stated LODs

Medium	Lab A	Lab B	Lab D
Ghost Wipe	N/A	0.0158 / 0.003-0.005	0.0139* /0.003-0.005
Whatman Filter	0.0139** / 0.005	N/A	0.0039 / 0.003-0.005

* High outlier omitted from computation
** Probable high outlier omitted from computation

Two conclusions are suggested by the results in Table 4. One is that laboratory D's LOD determination (0.003 - 0.005) is appropriate for Whatman Filter samples, but not for Ghost Wipes. The other is that the appropriate reporting limit for Ghost Wipes for labs B and D, and probably lab A as well, should be in the neighborhood of 0.015 - 0.020, regardless of the lab's reported LOD, in order for the reporting limit to have the meaning originally intended for L_C in Currie's system of decision, detection, and quantitation limits. Currie's <u>detection limit</u> and <u>quantiation limit</u> would clearly be even higher than those values.

One should note that Currie's L_C cleanly embodies the original concept of a "detected" analyte. However, one should also note that the Currie system of critical value, detection limit, and quantitation limit are intended to apply to the interpretation and use of individual measurements. In performing Be field characterization studies, our goal is most often to make credible decisions about the 95th percentile of the distribution of measurements; consequently, the selection of a "reporting limit" at which to censor data, if any, should be selected to facilitate inference about the 95th percentile as much as possible.

Why are the Blank Measurement Distributions Different?

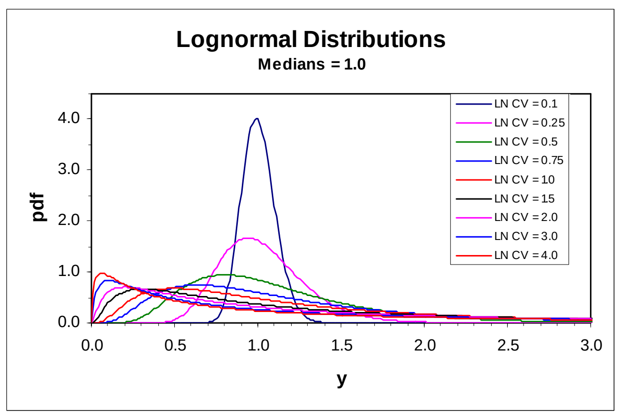
This is a great question; unfortunately, one rarely knows all of the details of how measurements are made, how calibrations are performed, etc. Some possibilities include the following:

- Differences in swipe media or swipe batches.
- Differences in how calibration standards are prepared (reagent water vs dissolved swipe media, e.g.).
- Differences in how calibration line is fit (forced through (0,0) or not).
- Differences in interferences inherent in selected emission line being reflected systematically in baselines.

Of these nearly all, except batch-to-batch variation in media, are controllable, but there appears to be a wide variability in practice.

Of particular interest is lab D using Whatman Filters, for which the fitted mean itself is negative. Similarly, a high proportion of samples obtained in office environments, including the buildings that triggered the initial NNSA investigation at North Las Vegas, gave negative values. In no case is the actual distribution of quantitated measurements restricted to positive numbers, except by censoring the data.

What About Overall Distributions of Be Measurements?

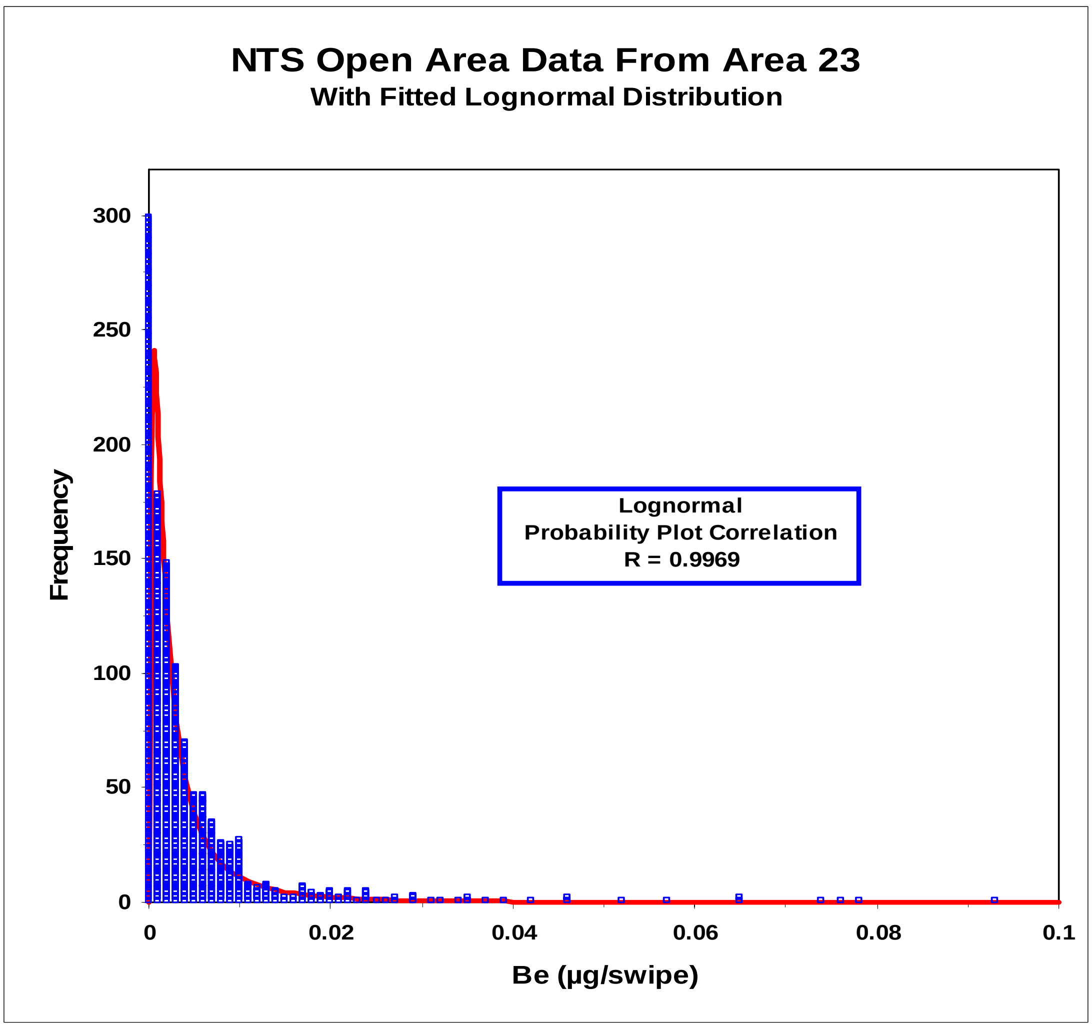


The assumption or assertion is often made that Be concentrations have non-negative lognormal distributions, such as those pictured here. Note that lognormal distributions range in shape from nearly normal to extremely skewed. While this may be a reasonable assumption about the actual <u>concentrations</u> themselves, it cannot hold for <u>measurements</u> because of the generally symmetric distributions of measurement error evidenced in the blank distributions and the occasionally substantial proportion of negative values.

Measurements should be thought of as resulting from a <u>variance components model</u> in which a measurement is the sum of the actual concentration, which may reasonably be thought of as lognormally distributed in many situations, and the normal distribution component of measurement variability; recall the symmetry of the distributions of the three scans that go into individual measurements.

Nonetheless, we feel that the lognormal model is a useful one for Be field characterization purposes, for the following reasons. So long as the reporting limit is sufficiently far below the regulatory threshold (0.2 µg/swipe), as will be the case with an adequate measurement system if Currie's L_C is used as the reporting limit or if the data are uncensored, the cases of interest will be those in which the spatial variability of actual concentrations dominates the mixture distribution. In those cases the measurement error will be a relatively smaller contribution to the overall distribution. If data are uncensored, of course, one will need to deal with the negative values in some fashion, possibly by thinking in terms of the three-parameter lognormal distribution family.

As an example, the following plot shows the overall distribution of measurements obtained in nearly all warehouses and similar facilities in Mercury on the NTS. These data were censored at 0.002 µg/swipe.



Summary of Suggestions

Regarding analytical methods:

- Prepare calibration standards in digestate of the swipe media of interest, in case there are any interferant present.
- Force calibration lines through (0,0), so that hopefully no more than around half the blanks have negative measurements.
- Use the 99% prediction limit from the distribution of blank quantitations as Currie's L_C; use this as the reporting limit IF one must have a reporting limit. But always remember that Be is naturally occurring, so that "finding" it is not a reportable event; the reportable event is not being able to show that the 95^{th} percentile is statistically significantly less than the DOE Release Criterion.
- AVOID THE 234.861 nm EMISSION LINE, because of the Fe interferant that is ubiquitous in the world, even in Ghost Wipes!

Regarding subsequent statistical analyses:

- If a reporting limit is needed, use Currie's L_C as discussed previously. In this case, further discussion of inference using censored data is given in the companion paper "Parametric 95%-95% Upper Tolerance Limits for Censored Lognormal Data" (C. Davis, 2005).
- Realize that the distribution of <u>measurements</u> is not lognormal; nonetheless, this is a reasonable working assumption for protection in the cases of most concern, with non-negligible spatial distributions of actual Be concentrations.

Acknowledgements:

Charles B. Davis, PhD	Principle Statistician, Envirostat Inc.
Dan Field, MS, CIH	Industrial Hygiene Program Manager, Bechtel Nevada
John Hess	IH Laboratory Manager, Bechtel Nevada
Dan Jensen	Industrial Hygienist II, Fluor Hanford PFP
Virginia Virgil	IH Laboratory Analyst, Bechtel Nevada

Disclaimer:
The results and ideas presented in this presentation paper reflect ongoing research work only, and do not represent the official position of Bechtel Nevada or the Department of Energy or any organization associated with the Beryllium Legacy Project at the Nevada Test Site.

Trademark Disclaimer:
Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors.